

Virtual Human Personality Masks: A Human Computation Approach to Modeling Verbal Personalities in Virtual Humans

Vaishnavi Krishnan¹, Adriana Foster², Regis Kopper¹, and Benjamin Lok¹

¹Computer and Information Sciences Engineering, University of Florida,
Gainesville, FL 32611, USA

{vakrishn, kopper, lok}@cise.ufl.edu

²Psychiatry and Health Behavior, Georgia Health Sciences University
Augusta, GA 30912, USA

afoster@georgiahealth.edu

Abstract. Modeling virtual humans that can exhibit realistic personalities is becoming increasingly important as virtual humans are being widely used for inter-personal skills education. We present Virtual Human Personality Masks, a system that combines human computation with the idea of using existing virtual humans to bootstrap the creation of other virtual humans to enable quick and easy generation of perceivable verbal personalities in virtual humans.

To evaluate this system, we created high and low verbosity-level variants of a virtual patient with symptoms of depression and conducted a user study with medical students split between two groups, each interacting with one of the two variants of the virtual patient. The participants' perceived verbosity levels of the virtual patients indicated that not only did the virtual patients created using our system exhibit the intended personality in a perceivable manner, but also exhibited other related personality attributes in a manner that is consistent with the human personality theory analogs of verbosity.

Keywords: Virtual Humans, Personalities, Verbosity, Human Computation, Conversational Agents, Virtual Patients.

1 Introduction

Virtual Humans (VHs) are increasingly being used as conversational partners in inter-personal skills education [1, 2]. To increase usefulness, inter-personal skills education should provide interactions with VHs that exhibit realistic personalities. In this paper, we propose a method for rapidly creating variants of existing VHs that can exhibit perceivable personalities using a human computation based approach[3].

Existing models of personality for conversational agents often use fixed sets of personality variables in combination with a decision algorithm to add verbal (the corpus) or non-verbal (gestures and emotions) personalities to the agent [4-9]. One such approach models the personality variables as states in a finite state machine [4]. Another approach uses a layered, Bayesian Belief Network to model personalities, moods and emotions [8]. Yet another approach uses multiple dimensions to represent a given personality state [9]. While these can be used to represent most generic personality states, the

amount of time required to apply such models to specific scenarios depends on the extensiveness of the model creation process and the amount of refining required to apply the model to the scenario. Our approach aims at reducing these dependencies using *personality masks* created out of real humans. The *personality masks* are created by gathering different human responses to a subset of the stimuli from an existing stimulus-response corpus and can then be applied onto the original corpus, resulting in different personality variants of the original VH. We built our system on top of Virtual People Factory (VPF), an online application for users to create and interact with VHS [11].

To demonstrate the generic concept of using the Personality Masks system to generate perceivable personalities, we created high and low *verbosity masks* and used the resulting *verbosity-masked* VHS to evaluate if the intended verbosity levels were perceivable to users. The results indicated that the VHS created using our system exhibit the intended verbosity levels in a manner that was perceivable for the interactants and consistent with the human personality theory analogs of verbosity.

2 Virtual Human Personality Masks

2.1 Verbosity and Patients with Depression

Patients' talkativeness(verbosity) levels in medical interviews can be influenced by factors such as their level of comfort with the interviewer, the environment, their age or the presence of personality disorders [12] and is often associated with extraversion, one of the five factors in the Big Five theory of personalities [13]. Extraverted people are perceived as easier to interview, more willing to disclose information and more cooperative and extraversion in patients is directly related to their perception of social support [14], suggesting that it may influence the patients' prognosis. This relevance of verbosity to the depression scenario and the fact that verbosity is easily perceivable and quantifiable, led us to choose it as the personality trait to be modeled for a 21-year old depressed VP, Cynthia Young (created using VPF) [15-18] using our system.

The two phases involved in the system are described in the following sections.

2.2 Mask Creation Phase

In this phase, human responses to a subset of the stimuli in the corpus are used to create different personality masks for the intended personality. To achieve this, we apply the findings of Rossen et al. [10] by using an existing question-answering VH (a VP) to create a question-asking VH (a virtual doctor), using the set of stimuli (what the users can say to the VH) and responses (what the VH will respond to the user) from the original corpus. The question-asking VH (virtual doctor) is then used to gather the personality-specific responses that will form the personality masks for the intended personality. Below, we describe the steps for the Mask Creation phase.

Step 1: Analysis of Interaction logs. Interaction logs of the existing corpus are analyzed to find the most frequently triggered stimuli. Since these stimuli are most likely to appear in future interactions with the VH, the ability of the VH to exhibit the intended personality while responding to these stimuli is pivotal to the design of our system. The threshold value for extracting such stimuli defaults to 25% usage, and is customizable by the domain expert. For the depression scenario, a set of approximately 100

stimulus-response pairs were filtered out from the original corpus with about 3500 stimuli and 400 responses.

Step 2: Review and refinement by domain expert. A domain expert reviews and refines the set of stimulus-response pairs filtered out in Step 1, using his/her expertise to determine which of the filtered stimuli can be used to elicit the intended personality trait. For example, the analysis of the interaction logs for our depressed VP corpus resulted in “What is your name?” as one of most frequently used stimuli, but is not useful in eliciting specifically long or short responses, not contributing to the creation of varied verbosity levels. For our example, 62 of the 100 most-frequently used stimulus-response pairs were selected by the domain expert for creating the virtual doctor.

Step 3: Response hint generation by domain expert. Once the set of interview questions are finalized, the domain expert fills-in response hints to guide the human interviewee to phrase their answers in a manner consistent with the original scenario. For example, the question “Have you had difficulty sleeping?” is open-ended and could be answered with excessive, normal or reduced sleep. However, in the original case history, the patient complained of excessive sleep and changing this key fact could completely change the diagnosis. The domain expert therefore added “Excessive sleep” as a response hint for that stimulus.

Step 4: Gathering responses from human interviewees. After the virtual doctor is created, an online chat-style interview link is sent out to several human interviewees who present varied verbosity levels. Before starting the interview, a description of the role that the interviewee will play during the interview is provided. In our example, this description asks the interviewee to role-play as a 21-year old student with symptoms of depression and talk to the virtual doctor as they would under such circumstances. During the interview, the virtual doctor asks questions to the interviewee while providing response hints for each question.

Step 5: Creation of Masks. Finally, the responses gathered from the human interviewees are cast into different *personality masks* by the domain expert using personality sliders that can be used to assign a “level” of the intended personality trait for each response. Fig. 1 shows an example of the verbosity slider for one of the 62 stimuli.

Question : Have you had difficulty sleeping?

Answer Hint : Excessive sleep

Available Responses :

☒ No, I am tired all the time. I can't seem to get myself around in the mornings and I sleep most of the day.

Very Chatty

☒ Not exactly difficulty, I actually sleep alot more than usual.

Somewhat Chatty

☒ yes, everynight

Neither Non-chatty nor Chatty

☒ Yes!

Very Non-chatty

Fig. 1. Verbosity sliders in the Personality Masks system for the depression example

2.3 Mask Application Phase

In this phase, the generated *personality masks* are applied onto the VH in the original corpus, resulting in a corpus that has *personality-masked* responses for some of the stimuli and default responses from the original corpus for the remaining stimuli.

3 Evaluation Study

We conducted a between-subjects user study with verbosity level of the VP as the independent factor. The participants were medical students (n=31, mean age = 24.2 years, 19 female and 12 male). They were randomly split into two groups and were assigned to interview either the high (n=16) or low (n=15) verbosity-level variant of the depressed VP in an online chat-style interaction. At the end of the interview, they completed a post-survey. Fig. 2 shows interaction transcripts for these conditions.

The two verbosity masks were created from the responses provided by 11 female staff and residents from the medical school for the virtual doctor’s questions in online chat-style interviews that lasted about 20 minutes each.

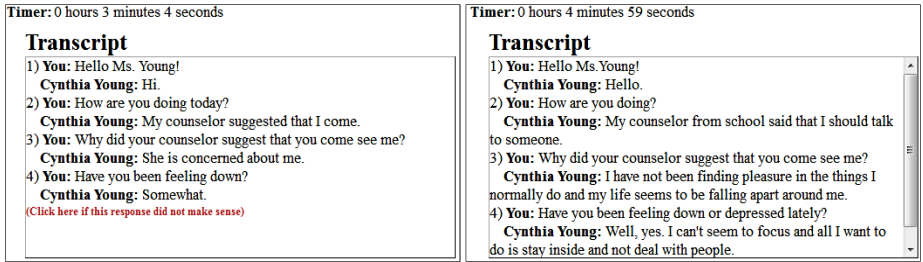


Fig. 2. Interaction transcripts of low (left) and high (right) verbosity VP

Primary Hypothesis. Participants in the low-verbosity group will perceive a lower level of verbosity in their VP than participants in the high-verbosity group.

Secondary Hypothesis. Participants in the low-verbosity group will perceive their VP as less concerned, less cooperating, less willing to disclose information and less comfortable to deal with than the participants in the high-verbosity group.

The metrics used in the post-survey for this study are summarized in Table 1.

Table 1. Likert scale rating (1-Strongly Disagree/5-Strongly Agree) statements for all metrics

Metric	Likert scale rating statements
Verbosity (Primary)	“The patient I spoke to gave detailed or long answers to the questions that I asked her in the interview”
Concern (Secondary)	“I think the patient was quite concerned about her problem”
Cooperation (Secondary)	“I think the patient cooperated well with me in answering my questions”
Self-disclosure (Secondary)	“I think the patient was willing to disclose information about herself.”
Comfort (Secondary)	“I liked talking to the patient and felt like I was able to help her.”

4 Results

The mean Likert scores for each metric across the two treatment groups are shown in Fig. 3. Table 2 summarizes the results of the Mann-Whitney U tests on the Likert scores. Results show that the scores for perceived levels of *verbosity*, *concern*,

cooperation, self-disclosure and comfort from participants in the low-verbosity treatment group were significantly lower than the scores for perceived verbosity-level from participants in the high-verbosity treatment group. All differences were significant at $p < 0.05$ or less. Thus, we **accept both our Hypotheses**.

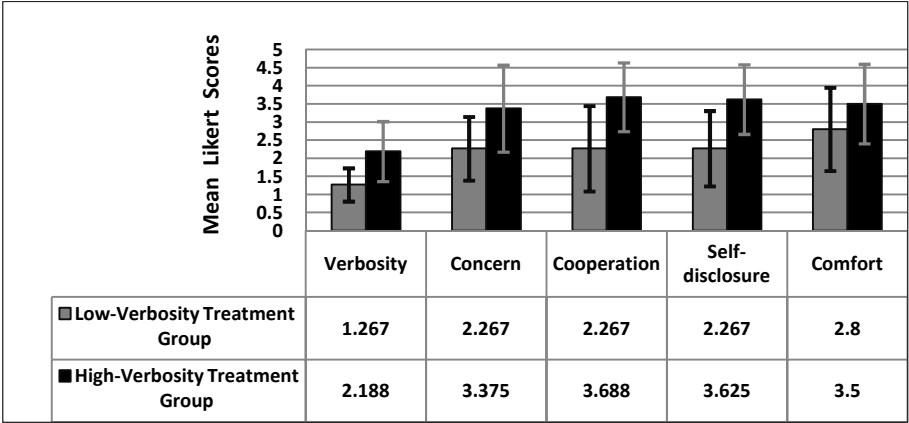


Fig. 3. Mean Likert Scores for all metrics (Error bars represent standard errors)

Table 2. Mann-Whitney U test statistics for all metrics

Metric	<i>U</i>	<i>Z</i>	<i>r</i>	<i>p</i> (one-tailed)
Verbosity	44.5	-3.24	0.58	0.001
Concern	59.0	-2.554	0.002	0.006
Cooperation	63.5	-2.325	0.42	0.010
Self-disclosure	61.5	-2.435	0.44	0.007
Comfort	79.5	-1.733	0.31	0.047

Table 3 summarizes the user comments, reinforcing the results from the post-survey.

Table 3. Examples of participants' responses to open-ended questions

Question Topic	High Verbosity Group	Low Verbosity Group
Overall experience	"It was a very accurate experience. I enjoyed the activity."	"It was very hard to ask further questions on a subject when she gave short or avoiding answers."
	"I felt like I wanted to speak more personally with her"	"Real patients are more forthcoming when you ask them questions such as 'why did you come here today?'"
Personality of the VP	"Very typical depressed mood, but eager to get back to feeling like herself."	"She seemed aware that there was a change in her behavior, but was hesitant to delve into the root of her issues."
	"She seemed very open & willing to talk to me even though she was depressed."	"Dry- perhaps she was simply depressed and down and having a hard time getting up the motivation to explain herself"

5 Discussion

Analysis of the results of our study indicates that participants were able to perceive the intended verbosity levels in the VPs they interviewed. The results for *concern*, *cooperation*, *self-disclosure* and *comfort* were also congruent with the Big Five theory of personality characteristics in which talkativeness and sociability are frequently associated with extraversion [13, 16]. These results also showed that our system has been successful in shifting the effort from modeling *standardized* personalities to modeling *perceivable* personalities.

However, we also observed that the participants' perceived verbosity-level of the high-verbosity VP were not as high as we expected them to be, even though the difference was statistically significant. One potential reason for this is the observation that the high-verbosity VP might not have been a considerable deviation from regular patients, as inferred from some of the previous work, that assert the fact that humans respond better to computer-based agents that are highly exaggerated [19].

6 Conclusions and Future Work

We have presented Virtual Human Personality Masks, a system that combines the concepts of human computation [3] with the idea of using existing VHS to bootstrap the creation of other VHS [10], to rapidly generate perceivable personalities in VHS. Below, we present the future directions for this research as identified by our experiences in designing the Personality Masks system and our understanding of potential applications for this system.

We aim to extend our work to model more complex personalities, such as the personality traits listed in the Big Five Personality model (OCEAN model) [15]. From our experiences with modeling verbosity as a personality trait for our depressed VP, we identified that modeling complex personalities would need inputs from other sources as well. In the case of our verbosity example, both the review and refinement of stimuli for the role-reversed VH and the response hint generation were done by the domain expert. However, for more complex personalities, psychology experts might provide guidelines on selecting the stimuli that can elicit the intended personality, while the domain experts provide response hints for each of those stimuli.

We also intend to explore the learning effects of using the *personality-masked* VHS in interpersonal skills education by running studies that measure the experience gained by interacting with VHS of different personalities.

We will also run follow-up studies to compare the extent of personality perception across different scenarios – for example, comparing the perception of personalities in the depression scenario (which needs a high-level of emotional connection between the patient and the doctor) with those in a Cranial Nerve disorder scenario (which mostly involves physical exams and less verbal interaction between the patient and the doctor). Through this study, we will aim to establish a relationship between the type of the scenario and the extent to which virtual human personalities would affect the learning of interpersonal skills in the scenario.

Acknowledgements. Thanks to Dr. Carolyn Stalvey, UF College of Medicine, for helping us find study participants, all participants for their time, reviewers for constructive

feedback and members/alumni of the Virtual Experiences Research Group, UF for their support.

References

1. Johnsen, K., et al.: The validity of a virtual human experience for interpersonal skills education. In: Proc. CHI 2007, pp. 1049–1058. ACM, San Jose (2007)
2. Randall, W.H., et al.: Virtual Humans in the Mission Rehearsal Exercise System (2003)
3. von Ahn, L.: Human computation. In: Proc. DAC 2009 (2009)
4. Badler, N.I., Reich, B.D., Webber, B.L.: Towards Personalities for Animated Agents with Reactive and Planning Behaviors. In: Creating Personalities for Synthetic Actors, Towards Autonomous Personality Agents, pp. 43–57. Springer (1997)
5. Ball, G., Breese, J.: Emotion and personality in a Conversational Character. In: Workshop on Embodied Conversational Characters, Tahoe City, CA (1998)
6. Doce, T., Dias, J., Prada, R., Paiva, A.: Creating Individual Agents through Personality Traits. In: Allbeck, J., Badler, N., Bickmore, T., Pelachaud, C., Safonova, A. (eds.) IVA 2010. LNCS, vol. 6356, pp. 257–264. Springer, Heidelberg (2010)
7. Kasap, Z., Magnenat-Thalmann, N.: Intelligent virtual humans with autonomy and personality: state of the art. In: Intelligent Decision Technologies, vol. 1, pp. 3–15 (2007)
8. Kshirsagar, S., Magnenat-Thalmann, N.: Virtual humans personified. In: Proceedings of the First International Joint Conference on Autonomous Agents and Multiagent Systems: Part 1, Bologna, Italy, pp. 356–357. ACM (2002)
9. Rushforth, M., Gandhe, S., Artstein, R., Roque, A., Ali, S., Whitman, N., Traum, D.: Varying Personality in Spoken Dialogue with a Virtual Human. In: Ruttkay, Z., Kipp, M., Nijholt, A., Vilhjálmsson, H.H. (eds.) IVA 2009. LNCS, vol. 5773, pp. 541–542. Springer, Heidelberg (2009)
10. Rossen, B., Cendan, J., Lok, B.: Using Virtual Humans to Bootstrap the Creation of Other Virtual Humans. In: Allbeck, J., Badler, N., Bickmore, T., Pelachaud, C., Safonova, A. (eds.) IVA 2010. LNCS, vol. 6356, pp. 392–398. Springer, Heidelberg (2010)
11. Rossen, B., Lind, S., Lok, B.: Human-Centered Distributed Conversational Modeling: Efficient Modeling of Robust Virtual Human Conversations. In: Ruttkay, Z., Kipp, M., Nijholt, A., Vilhjálmsson, H.H. (eds.) IVA 2009. LNCS, vol. 5773, pp. 474–481. Springer, Heidelberg (2009)
12. Sadock, B.J., Sadock, V.A.: Psychiatric History and Mental Status. In: Kaplan and Sadock's Synopsis of Psychiatry (2007)
13. Flesson, W., Gallagher, P.: The implications of Big Five standing for the distribution of trait manifestation in behavior: Fifteen experience-sampling studies and a meta-analysis. *J. Pers Soc. Psychol.* 97(6), 1097–1114 (2009)
14. Leskelä, U., et al.: The Influence of Major Depressive Disorder on Objective and Subjective Social Support: A Prospective Study. *J. Nerv. Ment. Dis.* 196(12), 876–883 (2008)
15. Goldberg, L.R.: An alternative description of personality: The big-five factor structure. *Journal of Personality and Social Psychology* 59, 1216–1229 (1990)
16. Nakao, K., et al.: The influences of family environment on personality traits. *Psychiatry and Clinical Neurosciences* 54(1), 91–95 (2000)
17. Hevil Shah, M.P.H., et al.: Interactive Virtual-Patient Scenarios: An Evolving Tool in Psychiatric Education. *Academic Psychiatry* 36, 146–150 (2012)
18. Nuzzarello, A.: C Birndorf, An Interviewing course for a psychiatry clerkship. *Academic Psychiatry* 28, 66–70 (2004)
19. Mateas, M.: An Oz-Centric Review of Interactive Drama and Believable Agents. In: Wooldridge, M.J., Veloso, M. (eds.) Artificial Intelligence Today. LNCS (LNAI), vol. 1600, pp. 297–328. Springer, Heidelberg (1999)